

# What does a large deviation look like?

David Andrieux

Large deviation theory quantifies the occurrence of events that deviate from the average behavior of a system. Such events arise from non-typical trajectories of the dynamics. In this note we derive the time evolution of these rare trajectories.

## I. PROBLEM FOMULATION

A large deviation calculation typically answers the question: What is the probability to observe, say, a value  $\xi$  of the entropy production? (or, more accurately, what is the exponential rate  $I(\xi)$  at which this probability decreases in time)

Here the questions we investigate are the following: *What does the dynamics leading to such a deviation  $\xi$  look like? Which trajectories are associated to such deviations?*

We address these questions in the context of Markov chains and derive the effective dynamics governing the large fluctuations of size  $\xi$ .

### A. Markov chains

We consider a Markov chain characterized by a transition matrix  $\hat{G} = (G_{ij}) \in \mathbb{R}^{N \times N}$  on the finite state space  $\Omega$ . The probability distribution  $\mathbf{p} = (p_1, p_2, \dots, p_N)$  evolves in discrete time steps according to

$$\mathbf{p}(n+1) = \mathbf{p}(n)\hat{G}. \quad (1)$$

The evolution operator  $\hat{G}$  is stochastic, i.e. it is non-negative ( $\hat{G} \geq 0$ ) and its rows sum to one ( $\sum_j G_{ij} = 1$ ). We assume that the Markov chain is primitive, i.e., there exists an  $n_0$  such that  $G^{n_0}$  has all positive entries. This guarantees that  $\hat{G}$  has a unique stationary distribution  $\boldsymbol{\pi}$  such that

$$\boldsymbol{\pi} = \boldsymbol{\pi}\hat{G}. \quad (2)$$

### B. Large deviations of the entropy production

For concreteness, we consider the large deviations of the entropy production. The entropy production is a fluctuating quantity measuring the dissipation occurring along a trajectory  $i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_k \rightarrow \dots \rightarrow i_n$  of the system. It takes the form

$$S(n) \equiv \ln \left( \prod_{k=0}^{n-1} \frac{G_{i_k i_{k+1}}}{G_{i_{k+1} i_k}} \right). \quad (3)$$

At long times, its probability distribution is characterized by the rate function

$$I(\xi) = \lim_{n \rightarrow \infty} -\frac{1}{n} \ln \text{Prob}[S(n)/n = \xi]. \quad (4)$$

An equivalent description is given in terms of the generating function

$$Q(\lambda) = \lim_{n \rightarrow \infty} -\frac{1}{n} \ln \langle e^{-\lambda S(n)} \rangle. \quad (5)$$

The generating function is related to the largest eigenvalue  $\rho(\lambda)$  of the operator

$$(\hat{L}_\lambda)_{ij} = G_{ij}^{1-\lambda} G_{ji}^\lambda \quad (6)$$

as follows:

$$Q(\lambda) = -\ln \rho(\lambda). \quad (7)$$

The operator (6) is non-negative but not stochastic:  $\sum_j (\hat{L}_\lambda)_{ij} \neq 1$  when  $\lambda \neq 0$ . When  $\lambda = 0$  we recover the original evolution operator:  $\hat{L}_{\lambda=0} = \hat{G}$ .

The descriptions in terms of the rate function  $I$  and the generating function  $Q$  are related through the Legendre transform

$$Q(\lambda) = \min_{\xi} [I(\xi) + \lambda \xi]. \quad (8)$$

These functions quantify the occurrence of events of size  $\xi$ . The goal of this paper is to go a step further and reveal their dynamical properties.

## II. LARGE DEVIATIONS ARE CHARACTERIZED BY AN EXTREMAL PRINCIPLE

We consider the space of stochastic matrices compatible with an irreducible non-negative matrix  $\hat{A} = (A_{ij}) \in \mathbb{R}^{N \times N}$ . A stochastic matrix  $\hat{P} = (P_{ij}) \in \mathbb{R}^{N \times N}$  is said to be compatible with  $\hat{A}$  if  $\hat{P}$  satisfies

$$1) \quad P_{ij} \geq 0 \quad \text{if} \quad A_{ij} \geq 0 \quad (9a)$$

$$2) \quad P_{ij} = 0 \quad \text{if} \quad A_{ij} = 0. \quad (9b)$$

We then have the following

**Theorem.** *Let  $\Sigma_{\hat{G}}$  be the space of stochastic matrices compatible with  $\hat{G}$ . Then*

$$Q(\lambda) = \min_{\Sigma_{\hat{G}}} \left[ \sum_{i,j} q_i P_{ij} \ln \left( \frac{P_{ij}}{G_{ij}} \right) + \lambda \sum_{i,j} q_i P_{ij} \ln \left( \frac{G_{ij}}{G_{ji}} \right) \right]. \quad (10)$$

Here  $\mathbf{q}$  is the stationary probability distribution of  $\hat{P}$  ( $\mathbf{q} = \mathbf{q} \hat{P}$ ).

DEMONSTRATION. Direct application of the Theorem 1.2 of Ref. [1] to the operator  $\hat{L}_\lambda$ .  $\square$

Furthermore, the optimal dynamics is given by the following

**Theorem.** *The equality in (10) holds for the stochastic dynamics*

$$\hat{P}_\lambda^* = \frac{1}{\rho(\lambda)} \text{diag}(\mathbf{x}_\lambda)^{-1} \hat{L}_\lambda \text{diag}(\mathbf{x}_\lambda) \quad (11)$$

and its stationary probability distribution

$$\mathbf{q}_\lambda^* = \frac{\mathbf{y}_\lambda \circ \mathbf{x}_\lambda}{\mathbf{y}_\lambda^T \mathbf{x}_\lambda}. \quad (12)$$

Here  $\mathbf{x}_\lambda > 0$  and  $\mathbf{y}_\lambda > 0$  are, respectively, the right and left eigenvectors of  $\hat{L}_\lambda$  corresponding to the largest eigenvalue  $\rho(\lambda)$ .  $\text{diag}(\mathbf{x})$  denotes the diagonal matrix with  $\mathbf{x}$  on its diagonal and  $\mathbf{y} \circ \mathbf{x}$  denotes the vector  $(y_1 x_1, \dots, y_N x_N)$ .

DEMONSTRATION: Insert (11) and (12) into (10) and check that we have equality.  $\square$

### III. INTERPRETATION

The result (10) has a clear physical interpretation. The term

$$D[\hat{P}] = \sum_{i,j} q_i P_{ij} \ln \left( \frac{G_{ij}}{G_{ji}} \right) \quad (13)$$

is the entropy production observed when fluctuations generate the statistics  $\hat{P}$ . The term

$$J[\hat{P}] = \sum_{i,j} q_i P_{ij} \ln \left( \frac{P_{ij}}{G_{ij}} \right) \quad (14)$$

is the Kullback-Leibler distance between the dynamics  $\hat{P}$  and  $\hat{G}$ . It corresponds to the probability rate to observe fluctuations with statistics  $\hat{P}$  in the original dynamics. The generating function at value  $\lambda$  thus selects out fluctuations that minimize the deviations from the original dynamics along with the  $\lambda$ -weighted entropy production.

The physical meaning of the terms (13) and (14) allows the connection with the rate function (4). The optimal dynamics  $\hat{P}_\lambda^*$  generates a dissipation  $\xi(\lambda) = D[\hat{P}_\lambda^*]$ . Inverting this relation, the rate function is given by

$$I(\xi) = J[\hat{P}_{\lambda(\xi)}^*]. \quad (15)$$

This establishes the link between the optimal Markov chains and the large deviations. The typical trajectories of the optimal chain  $\hat{P}_\lambda^*$  1) lead to an entropy production  $\xi(\lambda)$  and 2) appear at the rate  $I(\xi)$  in the original dynamics. We thus conclude that the large fluctuations of size  $\xi(\lambda)$ , when observed, appear through the dynamics  $P_{\lambda(\xi)}^*$ . This answers our original question on what large deviations look like.

### IV. TIME REVERSAL SYMMETRY AND LARGE DEVIATION PROPERTIES

The symmetry under time reversal implies a

**Fluctuation Theorem.** *The generating function has the symmetry*

$$Q(\lambda) = Q(1 - \lambda). \quad (16)$$

DEMONSTRATION: Direct consequence of the relation  $\hat{L}_\lambda = \hat{L}_{1-\lambda}^T$ .  $\square$

How does this symmetry affect the optimal chains? We start by looking at particular cases:

- $\lambda = 0$ : The optimal dynamics is identical to the original chain,  $\hat{P}_{\lambda=0}^* = \hat{G}$ , and the generating function  $Q(\lambda = 0) = 0$ .
- $\lambda = 1/2$ : The optimal dynamics  $P_{\lambda=1/2}^*$  satisfies the detailed balance conditions, leading to  $D = \xi = 0$ . The generating function thus takes the value  $Q(1/2) = I(\xi = 0)$ . That is, the optimal dynamics minimizes the Kullback-Leibler distance with respect to the original chain, given the detailed balance constraints.
- $\lambda = 1$ : The optimal dynamics is the time reversal of the original chain,  $\hat{P}_{\lambda=1}^* = \hat{G}^R$ . The time reversal of a Markov chain  $\hat{G}$  is defined as

$$\hat{G}^R = \text{diag}(\boldsymbol{\pi})^{-1} \hat{G}^T \text{diag}(\boldsymbol{\pi}) \quad (17)$$

where  $\boldsymbol{\pi} > 0$  is the stationary distribution of the chain  $\hat{G}$ .

More generally, we have the following

**Theorem.** *The chains  $\hat{P}_\lambda^*$  and  $\hat{P}_{1-\lambda}^*$  are time reversal of each other.*

DEMONSTRATION: Left as an exercise.

This theorem reveals that the paths leading to a given large deviation and those leading to its opposite one are time reversal of each other.

## V. EXAMPLE: LARGE DEVIATIONS OF A THREE-STATE SYSTEM

The generating function is given by  $Q(\lambda) = J(\lambda) + \lambda D(\lambda)$ , where  $D(\lambda) = D[\hat{P}_\lambda^*] = \xi(\lambda)$  and  $J(\lambda) = J[\hat{P}_\lambda^*]$  (Fig. 1). It satisfies the fluctuation theorem (section IV). The effective entropy production  $D(\lambda)$  is antisymmetric with respect to  $\lambda = 1/2$ , where it vanishes due to the equilibrium dynamics (section IV). At  $\lambda = 0$  it takes the value of the average entropy production of the original dynamics  $\hat{G}$ . The rate function  $I(\xi)$  is obtained from  $J(\lambda(\xi))$  (not shown).

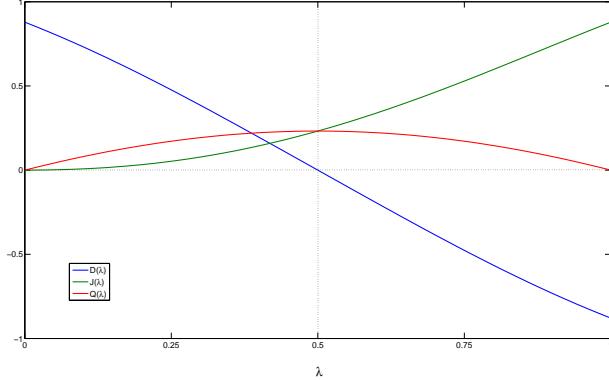


FIG. 1: **Generating function and its decomposition in a three-state system.** The generating function is determined by the rate  $J$  and the effective entropy production  $D$  of the optimal dynamics  $\hat{P}_\lambda^*$ . The transition probabilities take the values  $P_{12} = 0.8$ ,  $P_{13} = 0.2$ ,  $P_{21} = 0.1$ ,  $P_{23} = 0.9$ ,  $P_{31} = 0.7$ ,  $P_{32} = 0.3$ , and 0 otherwise.

Figure 2 illustrates the dynamics leading to different rare events, as measured by different values of  $\lambda$  or, equivalently, of the entropy production  $\xi(\lambda)$ . The first row corresponds to the original dynamics  $\hat{G}$ . There is a clear temporal structure in the trajectory, with an average positive circulation (i.e., a tendency to go in the direction  $1 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow \dots$ ). This average tendency disappears progressively until we reach  $\lambda = 1/2$ , which corresponds to an equilibrium dynamics and fluctuations of vanishing entropy production. For  $\lambda > 1/2$ , the fluctuations are, on average, going in the opposite direction (i.e., in the direction  $1 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow \dots$ ). Finally, the dynamics at  $\lambda = 1$  is the time reversal of  $\hat{G}$ . More generally, large deviations of opposite values are generated through a time-reversed dynamics (section IV).

## VI. CONCLUSIONS

Large deviations can be described in terms of an extremal principle over the space of Markov chains. The associated optimal chains describe the dynamical behavior of these large deviations. This approach presents neat mathematical properties and offers new insights into the dynamical origin of large deviations. The present results readily translate to other quantifiers of interest such as the thermodynamic currents. I believe that the knowledge of the paths leading to rare events will play an important role in the description and understanding of nonequilibrium systems.

**Disclaimer.** This paper is not intended for journal publication. Therefore, it lacks the delusional claims typically found in published papers about its (simultaneous) relevance to drug design, nanoelectronics, photonics, spintronics or any other fashionable application. Also, everyone can sleep at peace without worrying about missed citations and their h-index (and without e-mailing me to complain about them).

---

[1] Shi-Feng Shieh, Entropy **13**, 2036 (2011).

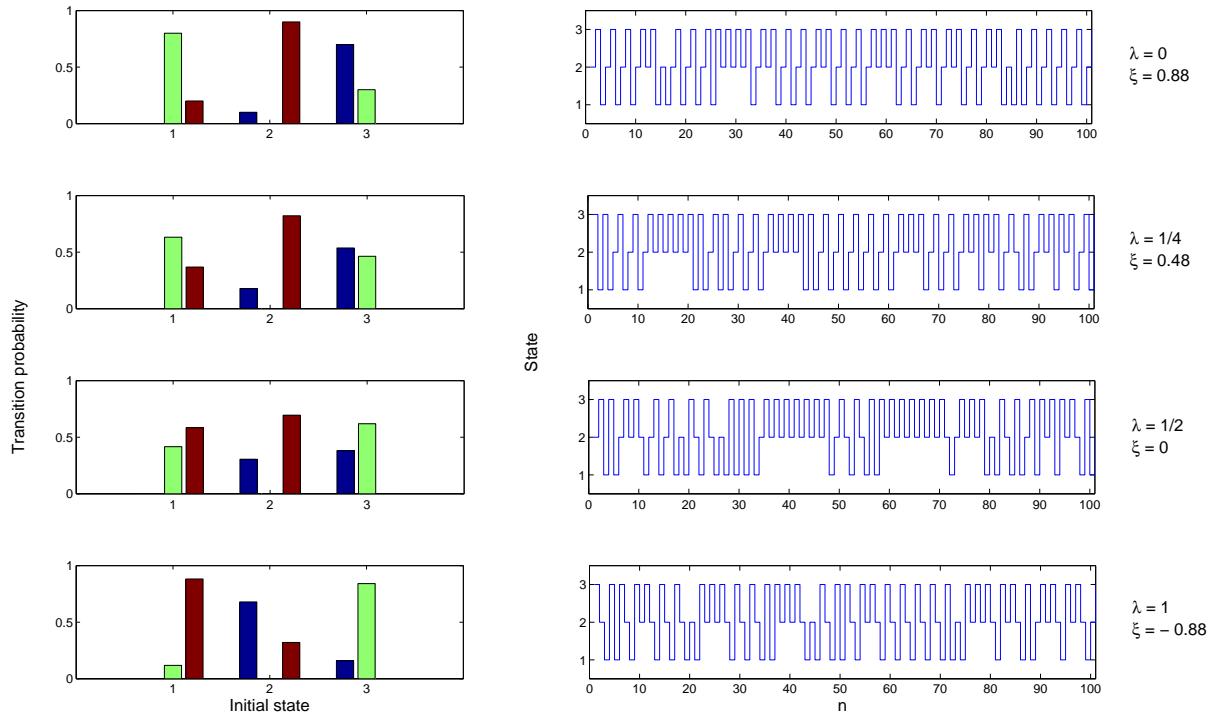


FIG. 2: **Large deviation dynamics in a three-state system.** Each row corresponds to a large deviation of given size (fixed  $\lambda$  or  $\xi$ ). The left-hand side displays the transition matrix of the optimal chains [final state = 1 (blue), 2 (green), 3 (red)]. The right-hand side shows the corresponding stochastic trajectories. The parameters take the same values as in Fig. 1.